REMARKS

Claims 1-23 are pending in this application. Claims 21-23, the title and drawings have been amended.

In paragraph 11 of the Office Action, the title of the invention is objected to. In response, Applicant has amended the title to be more descriptive of the invention.

In paragraph 12 of the Office Action, Fig. 5 is objected to as requiring the label "Prior Art." In response, Applicant has amended Figs. 5 and 6 to add the label "Prior Art."

In paragraph 14 of the Office Action, claims 1-23 are rejected under 35 U.S.C. § 112, first paragraph, as failing to comply with the enablement requirement. In particular, the Office Action asserts that training data must be used to generate models, and must have "training data to develop correct hypothesis." Applicant respectfully disagrees that the claims are not enabled and that training data must be used.

According to the present invention, the system generates models, initializes model parameters and estimates model parameters by referring only to an input text document as a processing target of text segmentation. *See, e.g.,* Specification at p. 13, l. 12 to p. 14, l. 26 (generating); p. 15, l. 8 to p. 16, l. 9 (initializing); and p. 16, l. 10 to p. 17, l. 27 (estimating). In the present invention, the vocabulary size can be determined also from an input text document. The vocabulary size, as a matter of course, may set to a sufficiently large constant value. No training data (or any data, other than the input text document) are used in the present invention. The specification, therefore, provides teachings that enable one with skill in the art to make and use the invention claimed. Therefore, Applicant respectfully asserts that the claims, as amended, are in full compliance with all § 112 requirements. Accordingly, withdrawal of this rejection, and the improper determination for the purposes of examination that "some sort of training data" is used, is respectfully requested.

In paragraph 13 of the Office Action, claim 22 is objected to as informal for containing an inappropriately placed period. In paragraph 15 of the Office Action, claims 21-23 are rejected under 35 U.S.C. § 112, second paragraph. In response, claims 21-23 have been amended to remove all informalities and to be in full compliance with all § 112

requirements. Accordingly, withdrawal of these objections/rejections is respectfully requested.

In paragraph 16 of the Office Action, claims 1, 3, 5-6, 10-12, 14, 16-17, 21 and 23 are rejected under 35 U.S.C. § 102(b) as being anticipated by U.S. Patent No. 6,104,989 (Kanevsky). Applicant traverses.

Among the limitations of independent claims 1, 10, 11, and 12 that are neither disclosed nor suggested in the prior art of record are the requirements for "generating a probability model in which information indicating each word of a text document belongs to a topic is made to correspond to a latent variable." On page 6, the Office Action alleges that Kanevsky discloses this feature at col. 2, ll. 25-28 and 45-53. Applicant disagrees.

Kanevsky discloses an approach based on an analogy from a voice recognition technique using Hidden Markov Modeling (HMM), in which a battery of topics are learned from processing training data. Kanevsky, col. 2, ll. 25-44. By using a large amount of learned data (other than the input text document), Kanevsky teaches that topics obtained from the learned data can correspond to latent variables. As taught by Kanevsky, "[t]hreshold ratios [obtained from the learned data] are established for determining [whether there is likely a] change of topics in the text." Kanevsky, col. 2, ll. 40-41. In contrast, claims 1, 10, 11 and 12 recite no such learning step of using training data nor do they require any training data at all.

An additional limitation among the limitations of independent claims 1, 10, 11 and 12 that are neither disclosed nor suggested in the prior art of record are the requirements for "outputting an initial value of a model parameter which defines the [generated] probability model." In a first embodiment comprising a discrete HMM, "parameters defining the model are state transition probabilities . . . and signal output probabilities." Specification, p. 15, ll. 12-14. Kanevsky does not generate a probability model that meets the limitations of claims 1, 10, 11 and 12, *i.e.*, a "model in which information indicating each word of a text document belongs to a topic is made to correspond to a latent variable." The Office Action cites Kanevsky at col. 5, lines 1-2 as supplying this teaching. Applicant disagrees.

Here at the cited portions, Kanevsky teaches that if "a conclusion [concerning emergence of a clear winner] is reached that a topic is not in the list [of topics identified in the training data], declare $\tilde{T}$ the current topic." The neutral topic $\tilde{T}$ is only a placeholder used to restart topic identification when no topic from the battery of topic can be identified. *See* col. 2, ll. 43-44. When a text string does not match any of the topics that were created in the battery by training data, this "neutral topic" is assigned. The neutral topic in Kanevsky segregates the text string to which it is assigned from that being translated unless more training data is used to add an additional topic to the battery. *See* col. 3, lns. 54-55. The selection of a neutral topic does not teach or suggest any initial values for the parameters that define a probability model. Nowhere within the four corners of Kanevsky are there any teachings for initializing model parameters. In the absence of any teaching or suggestion of this claimed feature of the invention, independent claims 1, 10, 11 and 12 are believed to be patentable over Kanevsky.

Among the limitations of independent claims 1, 10, 11, 12 and 21-23 that are neither disclosed nor suggested in the prior art of record are the requirements for "estimating a model parameter corresponding to a text document as a processing target on the basis of the initial value of the model parameter . . . and the text document." As discussed above, Kanevsky does not disclose or suggest initializing model parameters, and likewise, Kanevsky does not disclose or suggest estimating them either.

On page 7, the Office Action alleges that Kanevsky at col. 4, l. 57 to col. 5, l. 5 discloses this feature. Applicant disagrees. Here, Kanevsky teaches finding whether a candidate topic T_i for which the likelihood of the text is maximal, and comparing it against closest competing topics. If the likelihood is higher than a given factor, then T_i is the clear winner. Again, determining which topic among many learned from training data does not provide disclosure of estimating model parameters based on the initial values and the text document, as required by the claims. In the absence of any teaching or suggestion of this claimed feature of the invention, independent claims 1, 10-12 and 21-23 are believed to be patentable.

Dependent claims 2-9 and 13-20 depend either directly or indirectly from independent claims 1 and 12, and incorporate by reference all of the limitations found therein, and therefore are allowable for the same reasons expressed above. In addition, each of these dependent claims includes additional limitations which, in combination with the limitations incorporated by reference, are neither disclosed nor suggested in the art of record, and therefore are further allowable. Accordingly, claims 2-9 and 13-20 are likewise patentable.

In view of the above, each of the presently pending claims in this application is believed to be in immediate condition for allowance. Accordingly, the Examiner is respectfully requested to withdraw the outstanding rejection of the claims and to pass this application to issue.

Applicant respectfully requests an interview with the Examiner. The Examiner is welcome to contact the undersigned at the number listed below.

Dated: January 29, 2009

Respectfully submitted,

By _____

Robert G. Gingher
   Registration No.: 45,755
DICKSTEIN SHAPIRO LLP
1177 Avenue of the Americas
New York, New York 10036-2714
(212) 277-6500
Attorneys for Applicant